

On the Limits of Privacy in Reputation Systems

Stefan Schiffner

stefan.schiffner@esat.kuleuven.be

Andreas Pashalidis

andreas.pashalidis@esat.kuleuven.be

Elmar Tischhauser

elmar.tischhauser@esat.kuleuven.be

K. U. Leuven, ESAT/SCD/COSIC and IBBT
Kasteelpark Arenberg 10
B-3001 Leuven-Heverlee, Belgium

ABSTRACT

This paper describes a formal model for multiple privacy notions that apply to reputation systems and shows that, for certain classes of systems, very strong privacy notions are unachievable. In particular, it is shown that, systems where a user's reputation depends exclusively on the ratings he received, necessarily leak information about the relationship between ratings and reputations. In contrast, systems where a user's reputation depends both on the received ratings, and on the ratings received by others, potentially hide all information about this relationship. The paper concludes with guidelines for the construction of reputation systems that have the potential to retain high levels of privacy.

Categories and Subject Descriptors

D.4.6 [Operating Systems]: Security and Protection—*information flow controls*

General Terms

Algorithms, Design, Security

Keywords

Reputation Systems, Privacy, Limits, Anonymity, Unlinkability

1. INTRODUCTION

Reputation systems enable strangers to establish trust based on the premise that past behaviour is a good predictor of future behaviour. While reputation is typically conveyed through direct interaction or word-of-mouth in the real world, in the online world 'experience reports' from users are typically collected by a centralised system. This centralised system, called the 'reputation provider', then computes a reputation value for the 'reputation subjects', i.e. the entities that are mentioned in these reports [1]. The reputation provider also provides an interface over which

anyone can query the reputation of reputation subjects; this enables two strangers, namely a querier and a reputation subject, to establish trust.

The relationship between reputation and trust is examined in [2], where also alternative ways to establish trust are presented. The availability of reputation data may also have an impact on business; Dellarocas provides a useful overview of works that examine this aspect [3]. Multiple reputation systems have been proposed, some of which have been implemented and deployed. Currently, the most widely used reputation system is perhaps the one used by Ebay's auction platform. This system, primarily used by potential buyers in order to decide whether or not to place a bid, enables anyone to examine a seller's past transactions and ratings received by buyers.

The use of a reputation system raises privacy concerns. For example, as noted by Bygrave [4], ratings and reputation values may be seen as personal data. Providing privacy guarantees to the participants of a reputation system has also been proposed as a countermeasure to the problem of bad mouthing [5]. Recently proposed reputation systems aim to address the arising privacy concerns. The exact level of privacy achieved by these systems remains, however, unclear and this is partly due to the fact that reputation values themselves may influence privacy levels.

The contributions of this paper are threefold. Firstly, it formalises privacy for reputation systems in a way that enables effective comparisons between different systems. Secondly, it divides reputation systems into classes based on important properties such as liveness and non-discrimination. This division enables a systematic examination of different system types. Thirdly, it proceeds with such an examination by showing that there exists a fundamental tradeoff between privacy and other desirable properties. In particular, it is shown that 'strong anonymity' is unachievable if the system is non-discriminatory and a user's reputation exclusively depends on the ratings received by that user. It is also shown that, despite the use of pseudonyms, 'lively' reputation systems leak information *beyond* who has been rated and who has not.

The remainder of this paper is organised as follows. The next section briefly surveys related work. Section 3 introduces our formal system model, and Section 4 presents some simple example reputation system in terms of our model. Section 5 introduces the adversary model and the privacy definitions. Section 6 examines different properties of reputation functions which are then used in Section 7 to present

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WPES'11, October 17, 2011, Chicago, Illinois, USA.

Copyright 2011 ACM 978-1-4503-1002-4/11/10 ...\$10.00.

our main results. Finally, Section 8 concludes with guidelines for the construction of reputation functions.

2. RELATED WORK

Reputation systems differ in the way they compute the actual reputation of subjects; the surveys in [6, 7, 8] compare different approaches from multiple viewpoints, including privacy. More concrete proposals for privacy-preserving reputation systems are provided in [9, 10, 11, 6, 12, 13]. However, the above works lack a common formal privacy model and, hence, it remains unclear how to effectively compare their privacy properties. In this paper, we define a formal model that enables more comprehensive comparisons in terms of privacy.

Pavlov et al. propose to use anonymous communication in order to protect the privacy of those who query the reputation of ratees [14]. In this paper, we instead focus on privacy of raters and ratees. Certain design guidelines that aim to protect the privacy of both raters and ratees have been proposed in the literature. In particular, while [13] proposes to limit the number of possible reputation values in order to hamper an adversary’s ability to re-identify otherwise anonymous ratees, [15] proposes to compute a ratee’s reputation only after multiple ratings for the ratee have been collected. Our findings lead to another important guideline. Namely, we find that making a ratee’s reputation dependent on the ratings received by *other* ratees, potentially increases everyone’s privacy level.

Cryptosystems can also be used for the improvement of reputation system privacy. Androulaki et al. and Steinbrecher, in particular, propose the use of transaction pseudonyms in order to avoid linkability between transactions [9, 16]. Convertible credentials [13], electronic cash [9, 11] and a modified DC-network [17] have been proposed for the protection of rater privacy. Finally, Kerschbaum proposes a provably secure reputation system [18]. Our modelling covers the above system types. Moreover, in contrast to the above works, we also focus on the privacy impact of the choice of reputation function, i.e. the logic used to compute concrete reputation values for ratees.

Our model is an adaptation of the treatment described in [19, 20], which was originally proposed for the analysis of anonymous communication networks.

3. SYSTEM MODEL

A reputation system, denoted by \mathcal{S} in the sequel, enables its users to rate each other, and provides access to a reputation query interface over which users may query the current reputation value of the users. A reputation system consists of an enumerable but not necessarily finite user identifier space U such that $|U| \geq 2$, a pseudonym space P such that $|P| \geq |U|$, a space of ratings M , a space of reputation values V , and a *probabilistic algorithm* ϕ that maps all sequences of the form $((u, u', m)_1, (u, u', m)_2 \dots, (u, u', m)_n) \in (U^2 \times M)^n$ to a function that maps users to reputation values. We call ϕ the ‘reputation function’ and use $\phi(\mathcal{H})(u)$ to denote the reputation value in V that the algorithm produces for a given user u on input the sequence \mathcal{H} . A reputation system also provides the following four algorithms.

- The *initialization protocol* Init takes as input a security parameter $k \in \mathbb{N}$ and produces any required crypto-

graphic material. It also sets the initial history \mathcal{H} to be the empty sequence.

- The *rating protocol* Rate takes as input a triple $(u, u', m) \in U^2 \times M$ and appends it to \mathcal{H} . The rating protocol may also produce some output. Invoking the rating protocol means that user u rates user u' with the rating m .
- The *pseudonym generation protocol* NewPseudo takes as input a user $u \in U$ and outputs a pseudonym $p \in P$.
- The *reputation query protocol* GetReps outputs, for each $u \in U$, the pair $(\text{NewPseudo}(u), v) \in P \times V$, where $\text{NewPseudo}(u)$ is used to generate a fresh pseudonym for user u and where $v = \phi(\mathcal{H})(u)$ represents u ’s current reputation. The order in which the pairs are output depends on \mathcal{S} .

Although in this paper we assume that the space of reputation values V is discrete, it is straight-forward to extend our results to systems with continuous reputation value spaces.

A couple of remarks are in order. Firstly, observe that the NewPseudo algorithm is used internally by GetReps , as described above. We describe NewPseudo separately from GetReps because the logic of generating pseudonyms typically is, and should be, decoupled from the logic of calculating reputation. Actual cryptographic reputation systems are likely to have NewPseudo protocols that require interaction between the reputation provider and the affected user. However, at the end of a NewPseudo protocol execution, a user obtains only *his own* pseudonym.

Secondly, GetReps yields a list of pseudonyms and associated reputations for all users in the system. This modelling step was necessary in order to cover systems that assign unpredictable pseudonyms to users — it is obviously not possible to query someone’s reputation without being able to refer to him. Systems that use static or otherwise predictable pseudonyms, are also covered since the querier can simply discard irrelevant output. Moreover, since there is nothing that would stop a querier from querying the reputation of all users, outputting everyone’s (instead of a single user’s) reputation should not affect any privacy guarantees.

4. EXAMPLES

This section describes some simple reputation systems in terms of the model above; we briefly revisit these examples in later sections for illustration purposes. In all examples $\text{NewPseudo}(u)$ returns a bitstring of constant length, chosen uniformly at random (and, hence, pseudonyms reveal no information about user identities). Moreover, the reputation query protocol in all examples outputs the (p, v) pairs in an order that is chosen independently and uniformly at random for every GetReps protocol execution.

EXAMPLE 1. *The set of ratings $M = \{m\}$ is the singleton set (i.e. only positive feedback is possible), $V = \{0, \dots, c'\}$, and ϕ maps each $u \in U$ to a ‘reputation category’ (which can be thought of a number of ‘stars’) such that the worst and best reputation categories are 0 and $c' \geq 1$ respectively, and where the user advances to the next category after having accumulated $c \geq 1$ additional ratings. That is,*

$$\phi(\cdot)(u) = \begin{cases} 0 & \text{if } 0 \leq s_u < c \\ 1 & \text{if } c \leq s_u < 2c \\ \dots, & \\ c' & \text{if } s_u \geq c' \cdot c \end{cases}$$

where s_u denotes the number of ratings received by u . Note that the system described in [9] is a special case of this example.

EXAMPLE 2. $M = \{m\}$ (singleton), reputation values are encoded as natural numbers, i.e. $V = \mathbb{N}$, and ϕ maps each $u \in U$, to a ‘reputation category’ which is computed based on a static parameter $c \in \mathbb{N}^+$ as

$$\phi(\cdot)(u) = \begin{cases} 0 & \text{if } u \text{ received from } 0 \text{ up to } c-1 \text{ ratings} \\ 1 & \text{if } u \text{ received from } c \text{ up to } 2c-1 \text{ ratings} \\ \dots & \end{cases}$$

Note that, for $c = 1$, u ’s reputation is the total number of ratings u received.

EXAMPLE 3. $M = \{m\}$, $V = \mathbb{Z}$ and ϕ maps each $u \in U \setminus \{\text{Alice}\}$ to the number of ratings u received, minus the average $\lfloor |\mathcal{H}|/|U| \rfloor$. Alice is mapped to the reputation value 0.

EXAMPLE 4. The set of ratings is $M = \{-1, 1\}$ (i.e. both positive and negative feedback is possible), $V = \mathbb{Z}$, and ϕ maps each $u \in U$, to a ‘reputation category’ which is computed based on a static parameter $c \in \mathbb{N}^+$ as

$$\phi(\cdot)(u) = \begin{cases} \dots & \\ -1 & \text{if } -c \leq s_u < 0 \\ 0 & \text{if } 0 \leq s_u < c \\ 1 & \text{if } c \leq s_u < 2c \\ \dots, & \end{cases}$$

where s_u denotes the sum of ratings received by u . Note that, for $c = 1$, u ’s reputation is simply s_u .

EXAMPLE 5. The set of ratings is $M = \{m\}$, and the set of reputation values is $V = \{1, \dots, |U|\}$. The reputation function ϕ proceeds as follows. First, it calculates, for each $u \in U$, the sum of ratings received by u . Then it ranks all users according to this sum; the user with the highest sum is mapped to reputation value 1, the user with second highest sum to value 2, and so on. Ties are broken randomly and such that, in the end, every user is assigned a unique reputation value from V .

Note that, with the exception of Example 5, all our examples feature a deterministic reputation function.

5. PRIVACY FOR REPUTATION SYSTEMS

This section defines privacy for reputation systems in the face of an adversary. Let $n = |\mathcal{H}|$ denote the number of times **Rate** is invoked in a given time period. The correspondence between **Rate** invocations and the set of raters and ratees is modelled as two functions $\sigma, \rho \in \mathfrak{F}$, where $\mathfrak{F} = \{f : \{1, \dots, n\} \mapsto U\}$ is the space of functions that map (the serial number of) each **Rate** invocation to the user identifier space. That is, if (u, u', \cdot) is the parameter triple of the i th **Rate** invocation, then, for all $i \in \{1, \dots, n\}$, $\sigma(i)$ returns u and $\rho(i)$ returns u' .

Following ideas from [19, 20], the privacy notions considered in this paper describe potentially different degrees to

which σ and ρ remain hidden from an adversary. The adversary’s goal is to identify σ and ρ , or some ‘interesting property’ of these functions, possibly with respect to only some subset of **Rate** invocations, through interaction with, or observation of, the system \mathcal{S} . In particular, we consider the following properties of a function $f \in F$ with respect to a subset $I \subseteq \{1, 2, \dots, n\}$ of invocation serial numbers, which may be of interest to an adversary. These particular properties were chosen because they divide the function space in an intuitive way, and may be easily deducible from reputation values. For example, a user’s reputation value may reveal to the adversary how many ratings this user has received.

$U_{f,I} = \{f(i) : i \in I\} \subset U$ denotes the *participant set*, i.e. the set of user identifiers that are associated with the elements in I .

$Q_{f,I} = \{(u, \#_{u_{f,I}}) : u \in U_{f,I}\}$, where $\#_{u_{f,I}} = |\{i \in I : f(i) = u\}| \in \{1, 2, \dots, |U_{f,I}|\}$, denotes *usage frequency set*, i.e. the collection of records that indicate how many elements correspond to each participant from I ’s participant set.

$P_{f,I} = \{I'_1, I'_2, \dots, I'_{|U_{f,I}|}\} \vdash I$ denotes the *linking relation*, i.e. the partition of I that is induced by f . That is, $P_{f,I}$ denotes the partition that divides I into non-overlapping subsets such that, for all $i, i' \in I'_j$, $f(i) = f(i')$. Note that $\bigcup_j I'_j = I$.

$C_{f,I} = \{(1, c_1), (2, c_2), \dots, (|I|, c_{|I|})\}$ where, for all $i \in \{1, 2, \dots, |I|\}$, $c_i = |\{I' \in P_{f,I} : |I'| = i\}|$ denotes the *cardinalities of equivalence classes* as induced by the linking relation. Thus, $C_{f,I}$ is the multi set of equivalence class sizes with respect to the linking relation $P_{f,I}$.

In order to formalize privacy notions, we define the following nine notions of function distinguishability [20]. These notions combine the function properties defined above, and are used, in Def. 5.2 below, to define the nine privacy notions of ‘strong anonymity’ (SA), ‘strong unlinkability with participation hiding’ (SUP), ‘strong unlinkability with usage hiding’ (SUU), ‘weak unlinkability with participation hiding’ (WUP), ‘weak unlinkability with usage hiding’ (WUU), ‘weak unlinkability’ (WU), ‘pseudonymity’ (PS), ‘anonymity’ (AN), and ‘weak anonymity’ (WA).

DEFINITION 5.1. Two functions $f_0, f_1 \in \mathfrak{F}$, $f_0 \neq f_1$, are said, with respect to a subset of invocations $I \subseteq \{1, 2, \dots, n\}$, to be

SA-ind.	in any case,
SUP-ind.	iff $ U_{f_0,I} = U_{f_1,I} $,
SUU-ind.	iff $U_{f_0,I} = U_{f_1,I}$,
WUP-ind.	iff $C_{f_0,I} = C_{f_1,I}$,
WUU-ind.	iff $U_{f_0,I} = U_{f_1,I}$ and $C_{f_0,I} = C_{f_1,I}$,
WU-ind.	iff $Q_{f_0,I} = Q_{f_1,I}$,
PS-ind.	iff $P_{f_0,I} = P_{f_1,I}$,
AN-ind.	iff $U_{f_0,I} = U_{f_1,I}$ and $P_{f_0,I} = P_{f_1,I}$, and
WA-ind.	iff $Q_{f_0,I} = Q_{f_1,I}$ and $P_{f_0,I} = P_{f_1,I}$.

The adversary, denoted by \mathcal{A} , adaptively controls the usage of \mathcal{S} . Its interaction with \mathcal{S} is modelled via queries in an experiment that a challenger arranges for \mathcal{A} . At the beginning of this game, the challenger sets up \mathcal{S} by executing the

Experiment $\mathbf{Exp}_{\mathcal{S}, \mathcal{A}}^{X-b}(k)$
 $g \leftarrow \mathcal{A}^{\text{rate}((\cdot, \cdot, \cdot), (\cdot, \cdot, \cdot)), \text{getreps}()}$
return $g == b$

Figure 1: Experiment $\mathbf{Exp}_{\mathcal{S}, \mathcal{A}}^{X-b}(k)$ where $b \in \{0, 1\}$, $X \in \{\text{SA}, \text{SUP}, \text{SUU}, \text{WUP}, \text{WUU}, \text{WU}, \text{PS}, \text{AN}, \text{WA}\}$, k is \mathcal{S} 's security parameter, and where the challenger aborts the experiment if \mathcal{A} breaks any of the restrictions implied by notion X .

Init algorithm. The challenger then selects a bit $b \in \{0, 1\}$, uniformly at random, and then offers the following interfaces to \mathcal{A} , through which the system can be controlled.

- **rate** $((\cdot, \cdot, \cdot), (\cdot, \cdot, \cdot))$: on input $(u_0, u'_0, m_0), (u_1, u'_1, m_1) \in (U^2 \times M)^2$, the challenger executes $\text{Rate}(u_b, u'_b, m_b)$ and forwards \mathcal{S} 's output, if any, to \mathcal{A} .
- **getreps** $()$: the challenger invokes GetReps and forwards \mathcal{S} 's output to \mathcal{A} .

\mathcal{A} may issue a number of queries over these interfaces and, at some point in time, outputs a guess bit $g \in \{0, 1\}$. We say that \mathcal{A} wins the experiment if and only if $g = b$, and its advantage is given by

$$\mathbf{Adv}_{\mathcal{S}, \mathcal{A}}(k) = \left| \Pr(\mathbf{Exp}_{\mathcal{S}, \mathcal{A}}^{X-0}(k) = 0) - \Pr(\mathbf{Exp}_{\mathcal{S}, \mathcal{A}}^{X-1}(k) = 0) \right|.$$

In order to describe the course of an experiment we introduce the following notation. Let n and λ denote the number of **rate** and **getreps** queries respectively, that \mathcal{A} has issued up to the point in time it outputs g in an $\mathbf{Exp}_{\mathcal{S}, \mathcal{A}}^{X-b}(k)$ experiment. We define the functions $\sigma_0, \sigma_1, \rho_0, \rho_1$ such that, for all $i \in \{1, 2, \dots, n\}$, $\sigma_0(i) = u_{0,i}$, $\sigma_1(i) = u_{1,i}$, $\rho_0(i) = u'_{0,i}$, $\rho_1(i) = u'_{1,i}$, where $((u_{0,i}, u'_{0,i}, m_{0,i}), (u_{1,i}, u'_{1,i}, m_{1,i}))$ is the parameter tuple of \mathcal{A} 's i th **rate** query. We further define the subsets of **rate** query invocation serial numbers

$$\begin{aligned} I_1 &= \{1, \dots, c_1\}, \\ I_2 &= \{1, \dots, c_2\}, \\ &\vdots \\ I_\lambda &= \{1, \dots, c_\lambda\}. \end{aligned}$$

where, for all $j \in \{1, \dots, \lambda\}$, c_j denotes the number of **rate** queries \mathcal{A} has issued up to the point in time it issues the j th **getRep** query.

We use an adapted version of the privacy definition from [20].

DEFINITION 5.2. A reputation system \mathcal{S} is said to computationally provide ‘rater- X ’, denoted \mathbf{S}/X (resp. ‘ratee- X ’, denoted \mathbf{R}/X) for some privacy notion $X \in \{\text{SA}, \text{SUP}, \text{SUU}, \text{WUP}, \text{WUU}, \text{WU}, \text{PS}, \text{AN}, \text{WA}\}$, if and only if

- \mathcal{A} is restricted to **rate** invocation sequences $(u_0, u'_0, \cdot), (u_1, u'_1, \cdot), \dots$ such that σ_0 and σ_1 (resp. ρ_0 and ρ_1) are X -indistinguishable with respect to all $I \in 2^{\{I_1, \dots, I_\lambda\}}$,
- $\mathbf{Adv}_{\mathcal{S}, \mathcal{A}}(k) \leq \epsilon(k)$ for a negligible function $\epsilon(k)$, and
- \mathcal{A} 's running time is polynomial in k .

\mathcal{S} is said to statistically provide notion \cdot/X if and only if the first two conditions apply. Finally, \mathcal{S} is said to unconditionally provide \cdot/X if and only if the first condition applies and $\mathbf{Adv}_{\mathcal{S}, \mathcal{A}}(k) = 0$.

Figure 2 shows the hierarchy of privacy notions that naturally arises from the above definition. For brevity, we neither consider adversaries that may corrupt users, nor adversaries that are more severely restricted. Such consideration leads to both stronger and weaker variants the above privacy notions, and deriving such variants is straight forward [20]. Note that our main results in Section 7 only make use of the notions SA and SUU.

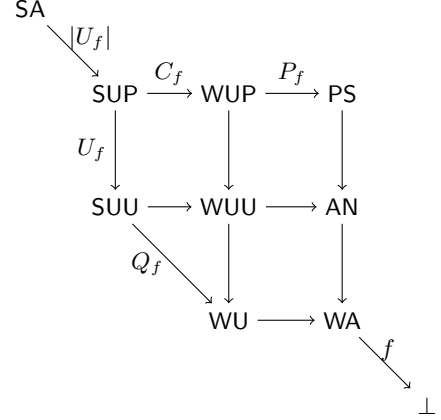


Figure 2: Relations between privacy notions, taken from [20]. The arrow labels indicate the property about f that the system may reveal.

REMARK 1. Example 1 with parameters $c = c' = 1$ provides \mathbf{R}/SUP and Example 2 with category parameter $c = 1$ provides \mathbf{R}/WUP . This is easy to see because, in the output of GetReps in Example 1, the number of reputation values that are equal to one is exactly the number of users that have received ratings, i.e. $|U_{\rho_b}|$. Similarly, in Example 2, reputation values reveal the number of ratings each user received, without revealing user identities; this is exactly the information encoded by C_{ρ_b} . Moreover, since neither pseudonym values, nor the order in which (p, v) pairs are output, reveal any information about Rate invocations, no information beyond $|U_{\rho_b}|$ (resp. C_{ρ_b}) is revealed. Example 5 provides \mathbf{R}/SA ; in order to see this, observe that the challenger’s output is independent from b : pseudonyms do not carry any information about incoming ratings, and the reputation values output by GetReps always (i.e. independently from the value of b) take the form of a uniformly at random chosen permutation of the sequence $(1, \dots, |U|)$.

REMARK 2. According to our system model (Section 3), incoming ratings (Rate protocol) always are associated with the user identifiers of both the rater and the ratee, while reputation values (GetReps protocol) are associated with fresh pseudonyms. This modelling covers systems where pseudonyms are used for ratings as well as for reputation outputs. In order to see this, observe that the domain of the functions ρ_0, ρ_1, σ_0 and σ_1 , over which privacy is defined, does not

represent pseudonyms or user identifiers, but instead system invocation serial numbers. Therefore, it does not matter whether pseudonyms or user identifiers are used in the modelling of the Rate protocol. Moreover, observe that systems that do not use pseudonyms or that use static pseudonyms are naturally supported, since the NewPseudo protocol may simply output the user’s identifier or a static pseudonym.

6. REPUTATION FUNCTIONS

This section examines different properties of reputation functions. Generally speaking, the reputation value reported by $\phi(\mathcal{H})(u)$ should enable one to obtain some information about the ratings that u received in the past, possibly in relation to the ratings that other users received. This, in turn, should enable an informed decision as to whether or not to enter a transaction with u . Intuitively, the more a user’s reputation value reveals about the received ratings, the higher the system’s utility. More precisely, a system with positive utility must report reputation values that depend on the history \mathcal{H} in some way. Moreover, utility increases with increasing dependence. For this reason, we use mutual information as a utility measure in this paper.

DEFINITION 6.1 (UTILITY). *The l -utility of a reputation function ϕ is defined as $T_{\phi,l} = I(\mathcal{H}, \phi(\mathcal{H}))$, where I denotes mutual information [21], and where \mathcal{H} is generated uniformly at random from the space of all histories of length $l \geq 1$. Moreover, the utility of a reputation function is defined as $T_{\phi} = \max_l T_{\phi,l}$. Furthermore, ϕ is said to have non-negligible utility if, for some $l \geq 1$, there exist two distinct histories $\mathcal{H}, \mathcal{H}'$ of length l and a function γ that maps users to reputation values, such that $\Pr(\phi(\mathcal{H}) = \gamma) > \Pr(\phi(\mathcal{H}') = \gamma) + \epsilon(k)$, where ϵ denotes a negligible function.*

Note that, although it cannot readily be mapped to meaningful usefulness notions in any given application, the above definition of utility captures all usefulness definitions that are based on the premise that *past behaviour influences future behaviour* in the sense that, if a reputation function is useful, then it must have positive utility according to our definition. Of course, for some entirely useless reputation functions ϕ , T_{ϕ} is positive; this, however, does not affect our results.

REMARK 3. *All examples from Section 4 have reputation functions with non-negligible utility. This is because, in all examples, the reputation value assigned to users depends on the ratings that are issued to the system.*

Some reputation systems can become ‘stuck’ in a state where new incoming ratings have no significant impact on anyone’s reputation. Whether or not a system can become stuck in such a state depends on the reputation function it uses. We call reputation functions, which ensure that the system *cannot* become stuck in such a state, ‘lively’.

DEFINITION 6.2 (LIVELINESS). *A reputation function ϕ is called weakly lively if and only if, for every \mathcal{H} there exists a \mathcal{H}' such that \mathcal{H} is a prefix of \mathcal{H}' and $\Pr(\phi(\mathcal{H}) = \gamma) = \Pr(\phi(\mathcal{H}') = \gamma')$, where γ and γ' are two distinct functions that map users to reputation values. Moreover, ϕ is called lively if and only if there exists some $v \in V$ such that $\Pr(\phi(\mathcal{H})(u) = v) \neq \Pr(\phi(\mathcal{H}')(u) = v)$ for all $u \in U$.*

REMARK 4. *Example 1 does not have a lively reputation function; indeed, when a user has received $c' - c$ ratings, his reputation switches from $c' - 1$ to c' , and, from this point onwards, remains ‘stuck’ at c' . Example 3 has a weakly lively reputation function because, while new ratings can generally influence user reputations, Alice’s reputation is stuck. All other examples from Section 4 have lively reputation functions, because new ratings potentially change the reputation value of any user.*

We distinguish reputation functions that, in order to compute a given user’s reputation, do *not* take into account the ratings received by other users. Such reputation functions are called ‘local’.

DEFINITION 6.3 (LOCALITY). *Let \mathcal{H}_u denote the subsequence of \mathcal{H} that arises when all entries where u does not appear as a ratee are removed. A reputation function ϕ is called local if and only if, for all \mathcal{H} , all $u \in U$, and all $v \in V$, $\Pr(\phi(\mathcal{H})(u) = v) = \Pr(\phi(\mathcal{H}_u)(u) = v)$.*

REMARK 5. *While Examples 1, 2 and 4 have local reputation functions, Examples 3 and 5 do not.*

Finally, we distinguish reputation functions that do not discriminate between different ratees, in the sense that they do not assign a better or worse reputation value to any user on the basis of his identity. We call such functions ‘non-discriminatory’. In order to define this formally, we first introduce the notion of history-user equivalences. Informally, two histories are equivalent with respect to two users, say Alice and Bob, if the two histories are identical except in that, whenever Alice was rated in the first history, Bob was rated in the second, and vice versa.

DEFINITION 6.4 (EQUIVALENT HISTORIES). *A quadruple $(\mathcal{H}, \mathcal{H}', u, u')$ comprised of two histories and two users is a history-user equivalence (alternatively, and the histories \mathcal{H} and \mathcal{H}' are equivalent with respect to u and u') if and only if \mathcal{H}' is the sequence that arises if all entries of \mathcal{H} of the form (\cdot, u, \cdot) are replaced with entries of the form (\cdot, u', \cdot) and \mathcal{H} is the sequence that arises if all entries of \mathcal{H}' of the form (\cdot, u', \cdot) are replaced with entries of the form (\cdot, u, \cdot) . The replacements preserve rater and rating values.*

DEFINITION 6.5 (NON-DISCRIMINATION). *A reputation function ϕ is called non-discriminatory if and only if, for all history-user equivalences $(\mathcal{H}, \mathcal{H}', u, u')$ and all $v \in V$, $\Pr(\phi(\mathcal{H})(u) = v) = \Pr(\phi(\mathcal{H}')(u') = v)$.*

Note that a non-discriminatory reputation function may still assign different weights to incoming ratings on the basis of raters identities. For example, a system that assigns greater weight to ratings issued by raters that have a good reputation than it assigns to ratings of worse-rated raters, may be non-discriminatory. Also note that, while liveliness and non-discrimination as well as a high utility are generally desirable properties for a reputation function its locality may or not be required.

REMARK 6. *The reputation function from Example 3 is obviously discriminatory. All other examples in Section 4 feature non-discriminatory reputation functions.*

7. RESULTS

In order to calculate a given user's reputation based on a reputation function that is both local and non-discriminatory, the identity of the user as well as the ratings received by other users are irrelevant; it is sufficient to consider only the ratings received by the user, possibly together with the identities of the corresponding raters. We formalise this rather intuitive observation by means of 'local histories', i.e. sequences that do not contain the identity of the ratee, as follows. Given a history \mathcal{H} , let $\tilde{\mathcal{H}}_u$ denote the sequence of the form $(u_1, m_1), (u_2, m_2), \dots, (u_{|\mathcal{H}_u|}, m_{|\mathcal{H}_u|})$ that arises when the ratee from each entry of \mathcal{H}_u is removed. Furthermore, let $\tilde{\phi}$ denote the probabilistic algorithm that, on input a sequence of the above form, outputs a reputation value; $\tilde{\phi}$ behaves as specified in Lemma 7.1, and is called ϕ 's local variant.

LEMMA 7.1. *For all local and non-discriminatory reputation functions ϕ there exists a probabilistic algorithm $\tilde{\phi}$ such that, for all \mathcal{H} , for all $u \in U$, and for all $v \in V$, $\Pr(\phi(\mathcal{H})(u) = v) = \Pr(\tilde{\phi}(\tilde{\mathcal{H}}_u) = v)$.*

PROOF. The result follows directly from Definitions 6.3 and 6.5. \square

LEMMA 7.2. *For all local and non-discriminatory reputation functions ϕ with $T_\phi > 0$ (resp. non-negligible utility), there exist two distinct local histories $\tilde{\mathcal{H}}, \tilde{\mathcal{H}}'$ and a reputation value $v \in V$ such that $\Pr(\tilde{\phi}(\tilde{\mathcal{H}}) = v) > \Pr(\tilde{\phi}(\tilde{\mathcal{H}}') = v)$ (resp. $\Pr(\tilde{\phi}(\tilde{\mathcal{H}}) = v) > \Pr(\tilde{\phi}(\tilde{\mathcal{H}}') = v) + \epsilon(k)$), where $\tilde{\phi}$ denotes ϕ 's local variant.*

PROOF. (Sketch) We show the case of positive utility. From Def. 6.1 follows that there exists some $l \geq 1$, two distinct histories $\mathcal{H}, \mathcal{H}'$ of length l , and a function $\gamma \in \Gamma$, where $\Gamma = \{\gamma : U \rightarrow V\}$, such that $\Pr(\phi(\mathcal{H}) = \gamma) > \Pr(\phi(\mathcal{H}') = \gamma)$. From Def. 6.3 follows that there exists some $u \in U$ such that \mathcal{H}_u and \mathcal{H}'_u are distinct and $\Pr(\phi(\mathcal{H}_u)(u) = v) > \Pr(\phi(\mathcal{H}'_u)(u) = v)$, where $v = \gamma(u)$. Moreover, from Lemma 7.1 follows that $\Pr(\tilde{\phi}(\tilde{\mathcal{H}}_u) = v) > \Pr(\tilde{\phi}(\tilde{\mathcal{H}}'_u) = v)$ and that, in fact, $\Pr(\tilde{\phi}(\tilde{\mathcal{H}}_{u'}) = v) > \Pr(\tilde{\phi}(\tilde{\mathcal{H}}'_{u'}) = v)$ for all $u' \in U$, where $\tilde{\phi}$ denotes ϕ 's local variant. The result follows since $\tilde{\mathcal{H}}_u$ and $\tilde{\mathcal{H}}'_u$ are distinct. The case of non-negligible utility is analogous. \square

We are now ready to present our main results. Theorems 7.3 and 7.4 below essentially state that, if a reputation function is non-discriminatory (i.e. a user's reputation does not depend on his identity) and local (i.e. a user's reputation depends on the ratings he received, and not on the ratings received by others), then any reputation system that employs this function leaks information about how individual ratings correspond to pseudonyms and their associated reputation values.

THEOREM 7.3. *No reputation system with at least two users and a local and non-discriminatory reputation function that has positive utility provides unconditional R/SA.*

PROOF. We describe an adversary \mathcal{A} with positive advantage in the SA-experiment. \mathcal{A} proceeds as follows. Based on ϕ 's description, it chooses two local histories $\tilde{\mathcal{H}}$ and $\tilde{\mathcal{H}}'$ and a value $v \in V$ such that

$$\alpha \neq \beta, \quad (1)$$

where $\alpha = \Pr(\tilde{\phi}(\tilde{\mathcal{H}}) = v)$, and $\beta = \Pr(\tilde{\phi}(\tilde{\mathcal{H}}') = v)$ where $\tilde{\phi}$ denotes ϕ 's local variant. We assume, without loss of generality that $|\tilde{\mathcal{H}}| \geq |\tilde{\mathcal{H}}'|$. It follows from Lemma 7.2 that the triple $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}', v)$ exists. \mathcal{A} then arbitrarily constructs another local history $\tilde{\mathcal{H}}''$ of length $|\tilde{\mathcal{H}}''| = |U||\tilde{\mathcal{H}}| - (|U|-1)|\tilde{\mathcal{H}}'|$ and selects a user $u \in U$, also arbitrarily, and starts an SA-experiment with the challenger where it issues $|U||\tilde{\mathcal{H}}|$ $\text{rate}(\cdot, \cdot)$ queries. In particular, in world $b = 0$, it rates all users with $\tilde{\mathcal{H}}$ and, in world $b = 1$, it rates all users in $U - \{u\}$ with $\tilde{\mathcal{H}}'$ and the user u with history $\tilde{\mathcal{H}}''$. Then it issues a getreps query and, as a result, obtains a (potentially randomly ordered) list of reputation values $(v_1, v_2, \dots, v_{|U|})$.

Let x denote the number of returned reputation values that are equal to v , $\alpha_\xi = \Pr(x = \xi \mid b = 0)$, and $\beta_\xi = \Pr(x = \xi \mid b = 1)$. The adversary, which can compute α_ξ and β_ξ for all $\xi \in \{0, 1, \dots, |U|\}$ based on ϕ 's description, outputs $g = 0$ if $\alpha_x > \beta_x$ and $g = 1$ otherwise.

We define the sets $\Xi = \{\xi : \xi \in \{0, 1, \dots, |U|\}, \alpha_\xi > \beta_\xi\}$ and $\Xi' = \{0, 1, \dots, |U|\} \setminus \Xi$. \mathcal{A} 's success probability $\Pr(g = b)$ is given by

$$\Pr(x \in \Xi \mid b = 0) \Pr(b = 0) + \Pr(x \in \Xi' \mid b = 1) \Pr(b = 1)$$

$$\begin{aligned} &= \frac{1}{2} \left[\sum_{\xi \in \Xi} \alpha_\xi + \sum_{\xi \in \Xi'} \beta_\xi \right] \\ &= \frac{1}{2} \left[1 - \sum_{\xi \in \Xi'} \alpha_\xi + \sum_{\xi \in \Xi'} \beta_\xi \right] = \frac{1}{2} + \frac{1}{2} \sum_{\xi \in \Xi'} (\beta_\xi - \alpha_\xi) \\ &= \frac{1}{2} \left[\sum_{\xi \in \Xi} \alpha_\xi + 1 - \sum_{\xi \in \Xi} \beta_\xi \right] = \frac{1}{2} + \frac{1}{2} \sum_{\xi \in \Xi} (\alpha_\xi - \beta_\xi). \end{aligned}$$

Since this means that \mathcal{A} has an advantage of

$$\mathbf{Adv}_{\mathcal{S}, \mathcal{A}}(k) = \frac{1}{2} \sum_{\xi \in \Xi'} (\beta_\xi - \alpha_\xi) = \frac{1}{2} \sum_{\xi \in \Xi} (\alpha_\xi - \beta_\xi),$$
 it follows

$$\text{that } 2 \cdot \mathbf{Adv}_{\mathcal{S}, \mathcal{A}}(k) = \frac{1}{2} \sum_{\xi \in \Xi'} (\beta_\xi - \alpha_\xi) + \frac{1}{2} \sum_{\xi \in \Xi} (\alpha_\xi - \beta_\xi)$$

$$\implies \mathbf{Adv}_{\mathcal{S}, \mathcal{A}}(k) = \frac{1}{4} \sum_{\xi=0}^{|U|} |\alpha_\xi - \beta_\xi|. \quad (2)$$

It must be shown that this sum is always positive, i.e. that, for some $\xi \in \{0, 1, \dots, |U|\}$, $\alpha_\xi \neq \beta_\xi$. Since α_ξ and β_ξ are probability distributions, it is sufficient to show that some of their moments are different. Let $\mu_{1, \alpha}$ and $\mu_{1, \beta}$ denote the expectation operator of the distributions, and $\mu_{2, \alpha}$ and $\mu_{2, \beta}$ their variances. Then we need to show that the system of equations

$$\begin{aligned} \mu_{1, \alpha} &= \mu_{1, \beta} \\ \mu_{2, \alpha} &= \mu_{2, \beta} \end{aligned}$$

has no solutions. From the fact that ϕ is local and non-discriminatory follows that

$$\alpha_\xi = \binom{n}{\xi} \alpha^\xi (1 - \alpha)^{n - \xi} \text{ and}$$

$$\beta_\xi = \binom{n-1}{\xi} \beta^\xi (1 - \beta)^{n - \xi - 1} (1 - \kappa) + \binom{n-1}{\xi-1} \beta^{\xi-1} (1 - \beta)^{n - \xi} \kappa.$$

It can be shown (see Lemma A.2 in Appendix A) that this leads to the system of equations

$$\begin{aligned} |U|\alpha &= (|U| - 1)\beta + \kappa \\ |U|\alpha(1 - \alpha) &= (|U| - 1)\beta(1 - \beta) + \kappa(1 - \kappa) \end{aligned}$$

which has solutions only if ($|U| = 1$ and $\alpha = \kappa$) or ($|U| > 1$ and $\alpha = \beta$). The first condition contradicts our assumption that $|U| \geq 2$, and the second our assumption that $\alpha \neq \beta$. \square

THEOREM 7.4. *No reputation system with a sufficiently large set of users and a local and non-discriminatory reputation function that has non-negligible utility provides statistical R/SA.*

PROOF. We show that the adversary described in the proof of Theorem 7.3 has non-negligible advantage. (We also reuse the notation introduced in that proof). Since the reputation function has non-negligible utility, it follows that $|\alpha - \beta| > \epsilon(k)$ where ϵ denotes a negligible function and k is the system's security parameter. We now examine the case where

$$\alpha > \beta + \epsilon(k). \quad (3)$$

For sufficiently large values of $|U|$, and if $0 \ll \alpha, \beta \ll 1$, the distributions of both α_ξ and β_ξ can be closely approximated by normal distributions \mathcal{N}_α and \mathcal{N}_β , respectively. (Recall that \mathcal{A} issues $n = |U||\tilde{\mathcal{H}}|$ Rate queries.) In the following, we use this approximation. We also assume that $\kappa = 1$, since this is the worst case because it maximises the similarity between the two distributions.

Suppose that the means $\mu_{1,\alpha}$ and $\mu_{1,\beta}$ of the two discrete probability distributions satisfy the condition

$$|\mu_{1,\alpha} - \mu_{1,\beta}| > \sigma_\alpha + \sigma_\beta, \quad (4)$$

where σ_α and σ_β denote their respective standard deviations. It is well-known that, for a variable y that follows a normal distribution, $\Pr(\mu - \sigma \leq y \leq \mu + \sigma) \approx 0.68$, and therefore also $\Pr(y < \mu - \sigma) \approx 0.16$. Similarly, $\Pr(y > \mu + \sigma) \approx 0.16$. From Eq. (2) follows that \mathcal{A} 's advantage is given by one fourth of the surface between the discrete distributions α_ξ and β_ξ and, hence, approximately also between their normal approximations. Moreover, Eq. (4) implies that the intervals $[\mu_{1,\alpha} - \sigma_\alpha, \mu_{1,\alpha} + \sigma_\alpha]$ and $[\mu_{1,\beta} - \sigma_\beta, \mu_{1,\beta} + \sigma_\beta]$ are non-overlapping. Hence,

Adv_{S,A}(k)

$$\begin{aligned} &= \frac{1}{4} \left(\Pr(\xi < \mu_{1,\alpha} + \sigma_\alpha \mid b = 0) - \Pr(\xi < \mu_{1,\alpha} - \sigma_\alpha \mid b = 1) \right. \\ &\quad \left. + \Pr(\xi > \mu_{1,\beta} - \sigma_\beta \mid b = 1) - \Pr(\xi > \mu_{1,\beta} + \sigma_\beta \mid b = 0) \right) \\ &\approx 1/4 \cdot 2 \cdot (0.68 + 0.16 - 0.16) = 0.34, \end{aligned}$$

which is clearly non-negligible. We now show that the condition of Eq. (4) holds if U is large enough. In particular, we show that Eq. (4) holds if

$$|U| > \delta^{-2}(1 - \beta) \left(2\beta + \delta + 2\sqrt{\beta^2 + \beta\delta} \right), \quad (5)$$

where $\delta = \alpha - \beta$. From Eq. (5) follows that

$$|\delta|U| + \beta - 1| > 2\sqrt{\beta(1 - \beta)}|U|. \quad (6)$$

Suppose that $\beta > 1/2 - \delta/2$.¹ From Eq. (6) follows that $1 - 2\beta - \delta < 0$ and that, hence, $\delta|U|(1 - 2\beta - \delta) < 0$. Therefore

$$\begin{aligned} &|\mu_{1,\alpha} - \mu_{1,\beta}| \\ &= \left| |U|(\beta + \delta) - ((n - 1)\beta + 1) \right| \\ &= \left| |U|\delta + \beta - 1 \right| \\ &> 2\sqrt{|U|\beta(1 - \beta)} \\ &= \sqrt{|U|\beta(1 - \beta)} + \sqrt{|U|\beta(1 - \beta)} \\ &> \sqrt{|U|\beta(1 - \beta)} + |U|\delta \cdot (1 - 2\beta - \delta) + \sqrt{(|U| - 1)\beta(1 - \beta)} \\ &= \sqrt{|U|(\beta + \delta)(1 - \beta - \delta)} + \sqrt{(n - 1)\beta(1 - \beta)} \\ &\geq \sqrt{|U|\alpha(1 - \alpha)} + \sqrt{(|U| - 1)\beta(1 - \beta) + \kappa(1 - \kappa)} \\ &= \sigma_\alpha + \sigma_\beta. \end{aligned}$$

This concludes the proof for the case where $\alpha > \beta + \epsilon(k)$. The case where $\beta > \alpha + \epsilon(k)$, and where therefore $\kappa = 0$ maximises the similarity between the two distributions, is handled analogously and is omitted. \square

Note that, as a trivial corollary of Theorem 7.4, it follows that no reputation system with a finite but sufficiently large set of users and a local and non-discriminatory reputation function that has non-negligible utility provides computational R/SA. Here, the only additional limitation is the finiteness of the user set; this limitation is required in order to ensure that the *computationally bounded* adversary is able to rate, and to process the reputation of, all users.

Finally, we show that reputation systems that are lively (in addition to being local and non-discriminatory), are more severely restricted. In particular, if a reputation function is lively (i.e. future ratings may influence a user's reputation), local, and non-discriminatory, then reputation systems employing this function they leak information beyond which pseudonyms correspond to users that have received ratings and which pseudonyms correspond to users that have not yet been rated.

THEOREM 7.5. *There exists no reputation system with a lively, local and non-discriminatory reputation function with positive utility that provides unconditional R/SUU.*

PROOF. We again use the adversary from the proof of Theorem 7.3. Based on ϕ 's description, this adversary chooses two local histories $\tilde{\mathcal{H}}$ and $\tilde{\mathcal{H}}'$ and a value $v \in V$ such that Eq. (3) holds. As in the proof of Theorem 7.3, the existence of these local histories follows from the fact that ϕ is both local and non-discriminatory. However, due to the liveness of ϕ , unlike in the proof of Theorem 7.3, both $\tilde{\mathcal{H}}$ and $\tilde{\mathcal{H}}'$ are non-empty. (In the proof of Theorem 7.3, $\tilde{\mathcal{H}}$ or $\tilde{\mathcal{H}}'$, but not both, may be empty.) Since the adversary rates all users in both worlds with non-empty histories, it follows that $U_{\rho_0} = U_{\rho_1} = U$. Therefore ρ_0 and ρ_1 are SUU-indistinguishable and, since the adversary has positive advantage, the result follows. \square

Theorem 7.5 holds also for statistical and computational R/SUU if the utility is non-negligible and the user set is finite and sufficiently large, respectively.

¹The case for $\beta \leq 1/2 - \delta/2$ can be shown analogously, and is omitted.

8. CONCLUSION AND FUTURE WORK

We presented a formal privacy model for reputation systems that places different notions into a well-characterised hierarchy. Based on this model, we showed that very strong privacy notions are unachievable for certain classes of reputation function. In particular, if a reputation function is non-discriminatory (i.e. a user's reputation does *not* depend on his identity) and local (i.e. a user's reputation depends on the ratings he received, and not on the ratings received by others), then any reputation system that employs this function does not achieve R/SA, i.e. leaks information about how individual ratings correspond to pseudonyms and their associated reputation values. Similarly, if a reputation function is *also* lively (i.e. future ratings may influence a user's reputation), then reputation systems employing this function do not even achieve the weaker notion R/SUU, i.e. they leak information beyond which pseudonyms correspond to users that have received ratings and which pseudonyms correspond to users that have not yet been rated.

We believe that liveliness and non-discrimination are very desirable properties for reputation functions, and do not recommend trading off these properties for privacy. However, our results suggest that non-local reputation functions, i.e. functions that, in order to compute a given user's reputation, do not only take into account the ratings received by the user, but also ratings received by other users, potentially lead to high privacy levels; we even described a trivial system (Example 5 in Section 4) that leaks no information about how ratings correspond to pseudonyms. This example is both lively and non-discriminatory.

Perhaps more surprisingly, it does not require the set of possible reputation values to be small. In fact, our analysis reveals that whether or not the number of possible reputation values has an impact on the achievable level of privacy depends on the details of the reputation function (see, for example, Example 2 with $c = c' = 1$ versus Example 3 with $c = 1$).

Non-local reputation functions seem to even have a utility advantage compared to local ones. This is because negative ratings not only degrade the reputation of the users whom these ratings are about, but also improve the reputation of the other users, and vice versa. In other words, non-local reputation functions typically require fewer ratings in order to adjust the 'overall picture', i.e. to update the *relative* reputation of all users. However, we also believe that utility is better measured in an application-dependent manner, and our rather technical definition does not necessarily lead to a meaningful utility score in the context of any given application.

An important future work item is the extension of our model with provisions for differential privacy [22]. We believe that a suitable extension is possible because rating histories can be treated like databases, the reputation function can be treated like a query algorithm that potentially adds noise for the purposes of sanitisation, and hence, and the adversary model can be adjusted accordingly. We expect that non-locality of the reputation function to be crucial under such an extension.

Our work raises multiple further research avenues. One obvious such avenue is the comparison of existing reputation systems based on our model. Such a comparison, which would not only encompass the reputation functions but also the protocol flows and data storage, can lead to the iden-

tification and removal of weaknesses. It is perhaps worth noting that a recent study, which used a similar model to compare RFID system, led to the identification of flaws [23].

Another potential future work topic is the examination of how privacy is affected by the number of allowed reputation values, as well as the number of required new ratings before a user's reputation is updated. Finally, a more detailed investigation of reputation function properties, for example based on quantitative metrics instead of a binary classification for properties such as non-discrimination and utility, is also an interesting research question. Such an investigation could lead to a toolbox with which one can systematically select reputation functions that achieve a desired tradeoff between privacy, utility, non-discrimination, and locality.

Acknowledgements

The authors are grateful to Sebastian Clauß, Vincent Rijmen, and Carmela Troncoso for their insightful comments and suggestions on an earlier version of this paper. The paper describes work undertaken partly in the context of the IAP Programme P6/26 BCRYPT of the Belgian State (Belgian Science Policy), partly in the context of the 'Trusted Architecture for Securely Shared Services' (TAS3) project (TAS3 is a collaborative project supported by the 7th European Framework Programme, with contract number 216287), and partly by the Research Council K.U. Leuven: GOA TENSE. Stefan Schiffner is partly supported by ENISA (www.enisa.europa.eu), and Elmar Tischhauser is a research assistant of the F.W.O., Fund for Scientific Research — Flanders.

9. REFERENCES

- [1] Resnick, P., Kuwabara, K., Zeckhauser, R., Friedman, E.: Reputation systems. *Communications of the ACM* **43**(12) (2000) 45–48
- [2] Artz, D., Gil, Y.: A survey of trust in computer science and the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web* **5**(2) (2007) 58 – 71 *Software Engineering and the Semantic Web*.
- [3] Dellarocas, C.: The digitization of word-of-mouth: Promise and challenges of online feedback mechanisms. *Management Science* (October 2003) 1407–1424
- [4] Bygrave, L.: *Data Protection Law, Approaching Its Rationale, Logic and Limits*. Kluwer Law International, The Hague, London, New York (2002)
- [5] Dellarocas, C.: Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In: *EC '00: Proceedings of the 2nd ACM conference on Electronic commerce*, New York, NY, USA, ACM Press (2000) 150–157
- [6] Voss, M.: Privacy preserving online reputation systems. In: *International Information Security Workshops*, Kluwer (2004) 245–260
- [7] Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. *Decision Support Systems* **43**(2) (2007) 618 – 644 *Emerging Issues in Collaborative Commerce*.
- [8] Mui, L.: *Computational Models of Trust and Reputation: Agents, Evolutionary Games, and Social Networks*. PhD Thesis, Massachusetts Institute of Technology (2003)

- [9] Androulaki, E., Choi, S.G., Bellovin, S.M., Malkin, T.: Reputation systems for anonymous networks. In: PETS '08: Proceedings of the 8th international symposium on Privacy Enhancing Technologies, Berlin, Heidelberg, Springer-Verlag (2008) 202–218
- [10] Voss, M., Heinemann, A., Mühlhäuser, M.: A Privacy Preserving Reputation System for Mobile Information Dissemination Networks. In: First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SECURECOMM'05), IEEE (2005) 171–181
- [11] Schiffner, S., Clauß, S., Steinbrecher, S.: Privacy and liveliness for reputation systems. In: Proceedings of 2009 European PKI Workshop (EuroPKI'09), Springer (2010) (to appear).
- [12] ENISA: Position paper. reputation-based systems: a security analysis. available from http://www.enisa.europa.eu/doc/pdf/deliverables/enisa_pp_reputation_based_system.pdf (last visit 16/06/09) (2007)
- [13] Steinbrecher, S.: Enhancing multilateral security in and by reputation systems. In: Proceedings of the IFIP/FIDIS Internet Security and Privacy Summer School, Masaryk University Brno, 1-7 September 2008. Volume 298 of IFIP AICT., Springer (2009) 135–150
- [14] Pavlov, E., Rosenschein, J.S., Topol, Z.: Supporting privacy in decentralized additive reputation systems. In: The Second International Conference on Trust Management, Oxford, United Kingdom (March 2004) 108–119
- [15] Dellarocas, C.: Research note – how often should reputation mechanisms update a trader's reputation profile? *Information Systems Research* **17**(3) (2006) 271–285
- [16] Steinbrecher, S.: Design options for privacy-respecting reputation systems within centralised internet communities. In: Proceedings of IFIP Sec 2006, 21st IFIP International Information Security Conference: Security and Privacy in Dynamic Environments. Volume 201 of IFIP., Springer (May 2006) 123–134
- [17] Schiffner, S., Clauß, S., Steinbrecher, S.: Fairness and Information-theoretic Privacy for Reputation. In Hromkovič, J., Královič, R., eds.: *SOFSEM 2011: 37th Conference on Current Trends in Theory and Practice of Informatics*. Volume 6543 of *Lecture Notes in Computer Science*, Nový Smokovec, SK, Springer-Verlag (2011) 16
- [18] Kerschbaum, F.: A verifiable, centralized, coercion-free reputation system. In: Proceedings of the 8th ACM workshop on Privacy in the electronic society. WPES '09, New York, NY, USA, ACM (2009) 61–70
- [19] Hevia, A., Micciancio, D.: An indistinguishability-based characterization of anonymous channels. In Borisov, N., Goldberg, I., eds.: *Privacy Enhancing Technologies*. Volume 5134 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg (2008) 24–43 10.1007/978-3-540-70630-4_3.
- [20] Bohli, J.M., Pashalidis, A. In: *Relations Among Privacy Notions*. Springer-Verlag, Berlin, Heidelberg (2009) 362–380
- [21] Cover, T.M., Thomas, J.A.: *Elements of Information Theory*. John Wiley & Sons Inc., Hoboken (2005)
- [22] Dwork, C.: Differential privacy. In Bugliesi, M., Preneel, B., Sassone, V., Wegener, I., eds.: *Automata, Languages and Programming*. Volume 4052 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg (2006) 1–12
- [23] Hermans, J., Pashalidis, A., Vercauteren, F., Preneel, B.: A New RFID Privacy Model. In: 2011st European Symposium on Research in Computer Security (ESORICS 2011). *Lecture Notes in Computer Science*, Leuven, BE, Springer-Verlag (2011) 20

APPENDIX

A. MOMENTS OF THE SKEWED BINOMIAL DISTRIBUTION

Consider two Bernoulli trials with success probabilities β and κ , respectively. Denote by X the random variable describing the number of combined successes if the first trial is repeated $n - 1$ times and the second is executed once. The combined probability of having ξ successes over all n trials is then given by

$$\Pr(X = \xi) = \binom{n-1}{\xi} \beta^\xi (1-\beta)^{n-\xi-1} (1-\kappa) + \binom{n-1}{\xi-1} \beta^{\xi-1} (1-\beta)^{n-\xi} \kappa.$$

We are now interested in closed formulas for the first and second moment of this distribution.

LEMMA A.1. *For all positive integers m ,*

$$\sum_{i=0}^n i^m \binom{n}{i} p^i (1-p)^{n-i} = np \sum_{i=0}^n i^{m-1} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i}. \quad (7)$$

PROOF. The results follows from the well-known fact that $k \binom{n}{k} = n \binom{n-1}{k-1}$ for $k > 0$. \square

LEMMA A.2. *The first and second moments of the distribution (A) are given by*

$$\mu_1 = (n-1)\beta + \kappa \quad \text{and} \quad (8)$$

$$\mu_2 = (n-1)\beta(1-\beta) + \kappa(1-\kappa). \quad (9)$$

PROOF. We have

$$\begin{aligned}
\mu_1 &= \sum_{i=0}^n i \cdot \Pr(X = i) \\
&= (1 - \kappa) \sum_{i=0}^n i \binom{n-1}{i} \beta^i (1 - \beta)^{n-i-1} \\
&\quad + \kappa \sum_{i=0}^n i \binom{n-1}{i-1} \beta^{i-1} (1 - \beta)^{n-i} \\
&= (1 - \kappa) \sum_{i=0}^{n-1} i \binom{n-1}{i} \beta^i (1 - \beta)^{n-i-1} \\
&\quad + \kappa \sum_{i=1}^n i \binom{n-1}{i-1} \beta^{i-1} (1 - \beta)^{n-i} \\
&= (1 - \kappa) \cdot (n-1)\beta \\
&\quad + \kappa \sum_{i=0}^{n-1} (i+1) \binom{n-1}{i} \beta^i (1 - \beta)^{n-i-1} \\
&= (1 - \kappa) \cdot (n-1)\beta \\
&\quad + \kappa \left(\sum_{i=0}^{n-1} i \binom{n-1}{i} \beta^i (1 - \beta)^{n-i-1} \right. \\
&\quad \quad \left. + \underbrace{\sum_{i=0}^{n-1} \binom{n-1}{i} \beta^i (1 - \beta)^{n-i-1}}_{=1} \right) \\
&= (1 - \kappa) \cdot (n-1)\beta + \kappa((n-1)\beta + 1) \\
&= (n-1)\beta + \kappa
\end{aligned}$$

and

$$\mu_2 = \sum_{i=0}^n i^2 \cdot \Pr(X = i) - \mu_1^2,$$

and this sum can be evaluated as

$$\begin{aligned}
&\sum_{i=0}^n i^2 \cdot \Pr(X = i) \\
&= (1 - \kappa) \sum_{i=0}^n i^2 \binom{n-1}{i} \beta^i (1 - \beta)^{n-i-1} \\
&\quad + \kappa \sum_{i=0}^n i^2 \binom{n-1}{i-1} \beta^{i-1} (1 - \beta)^{n-i} \\
&= (1 - \kappa) \sum_{i=0}^{n-1} i^2 \binom{n-1}{i} \beta^i (1 - \beta)^{n-i-1} \\
&\quad + \kappa \sum_{i=0}^{n-1} (i+1)^2 \binom{n-1}{i} \beta^i (1 - \beta)^{n-i-1}
\end{aligned}$$

which can be reformulated by Lemma A.1 to

$$\begin{aligned}
&= (1 - \kappa) \cdot (n-1)\beta \sum_{i=0}^{n-1} i \binom{n-2}{i-1} \beta^{i-1} (1 - \beta)^{n-i-1} \\
&\quad + \kappa \cdot (n-1)\beta \cdot \left(\sum_{i=0}^{n-1} i \binom{n-2}{i-1} \beta^{i-1} (1 - \beta)^{n-i-1} \right. \\
&\quad \quad \left. + 2 \sum_{i=0}^{n-1} i \binom{n-1}{i} \beta^i (1 - \beta)^{n-i-1} \right. \\
&\quad \quad \left. + \sum_{i=0}^{n-1} \binom{n-1}{i} \beta^i (1 - \beta)^{n-i-1} \right) \\
&= (1 - \kappa)(n-1)\beta \sum_{i=0}^{n-2} (i+1) \binom{n-2}{i} \beta^i (1 - \beta)^{n-i-2} \\
&\quad + \kappa \cdot (n-1)\beta \cdot \left(\sum_{i=0}^{n-1} i \binom{n-2}{i-1} \beta^{i-1} (1 - \beta)^{n-i-1} \right. \\
&\quad \quad \left. + 2(n-1)\beta + 1 \right) \\
&= (1 - \kappa)(n-1)\beta((n-2)\beta + 1) \\
&\quad + \kappa \cdot (n-1)\beta \cdot \left(\sum_{i=0}^{n-2} (i+1) \binom{n-2}{i} \beta^i (1 - \beta)^{n-i-2} \right. \\
&\quad \quad \left. + 2(n-1)\beta + 1 \right) \\
&= (1 - \kappa)(n-1)\beta((n-2)\beta + 1) \\
&\quad + \kappa((n-1)\beta((n-2)\beta + 3) + 1) \\
&= (n-1)\beta((n-2)\beta + 2\kappa + 1) + \kappa \\
&= (n-1)\beta((n-2)\beta + 2\kappa + 1) + \kappa - ((n-1)\beta + \kappa)^2
\end{aligned}$$

and hence $\mu_2 = (n-1)\beta(1 - \beta) + \kappa(1 - \kappa)$. \square